DOCUMENT RESUME

ED 477 080 IR 021 808

AUTHOR	Park, Young; Bauer, Malcolm
TITLE	The Role of a Reusable Assessment Framework in Designing Computer-Based Learning Environments.
PUB DATE	2002-06-00
NOTE	7p.; In: ED-MEDIA 2002 World Conference on Educational Multimedia, Hypermedia & Telecommunications. Proceedings (14th, Denver, Colorado, June 24-29, 2002); see IR 021 687.
AVAILABLE FROM	Association for the Advancement of Computing in Education (AACE), P.O. Box 3728, Norfolk, VA 23514. Tel: 757-623-7588; e-mail: info@aace.org; Web site: http://www.aace.org/DL/.
PUB TYPE	Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE	EDRS Price MF01/PC01 Plus Postage.
DESCRIPTORS	*Computer Assisted Instruction; Computer System Design; *Curriculum Design; Design Preferences; *Educational Assessment; Educational Technology; *Evaluation Methods; *Instructional Design; Instructional Development
IDENTIFIERS	*Design Methodology

ABSTRACT

This paper introduces the concept of a reusable assessment framework (RAF). An RAF contains a library of linked assessment design objects that express: (1) specific set of proficiencies (i.e. the knowledge, skills, and abilities of students for a given content or skill area); (2) the types of evidence that can be used to estimate those proficiencies; and (3) features of tasks that will aid in the design of activities (e.g. features that need to be present in order for students to produce the evidence, features that affect task difficulty, etc.). While RAFs can speed the design of many kinds of assessments, in this paper the focus is on their use to aid instructional designers in embedding assessments within computer-based learning environments. The RAF concept is based upon the evidence-centered design methodology described in Mislevy, Steinberg, Almond, Haertel, & Penuel (2001). (Author)

Reproductions supplied by EDRS are the best that can be made
from the original document.

The Role of a Reusable Assessment Framework in Designing Computer-Based Learning Environments

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

G.H. Marks

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

Young Park (yp85@columbia.edu),
Teachers College, Columbia University
Malcolm Bauer (mbauer@ets.org),
Educational Testing Service

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Abstract: This paper will introduce the concept of a reusable assessment framework (RAF). An RAF contains a library of linked assessment design objects that express a) specific set of proficiencies (i.e. the knowledge, skills, and abilities of students for a given content or skill area), b) the types of evidence that can be used to estimate those proficiencies, and c) features of tasks that will aid in the design of activities (e.g. features that need to be present in order for students to produce the evidence, features that affect task difficulty, etc.). While RAFs can speed the design of many kinds of assessments, in this paper we focus on their use to aid instructional designers in embedding assessments within computer-based learning environments. The RAF concept is based upon the evidence-centered design methodology described in Mislevy, Steinberg, Almond, Haertel, & Penuel (2001).

ED 477 080

The Problem

There has been substantial discussion and research concerning the advantages of embedding assessments within instructional activities and learning environments (e.g. Shavelson, 1992; Hopper, 1998; Treagust, Jacobowitz, Gallagher, and Parker, 2000; Van Lehn, 1988). Van Lehn (1988), for example, describes several ways in which embedded assessments can help people learn including a) determining appropriate tasks based upon estimates of students' skill level, b) determining when to provide feedback, and c) adapting explanations to students' level of understanding. However, designing learning environments that embed assessments is a labor-intensive activity. There are several reasons for this. First, creating learning environments is already hard. Given the time and budget constraints that instructional designers face, it is hard to justify the added time effort, time, and expense of incorporating reliable, valid assessments as part of a learning environment. Secondly, designing embedded assessments requires an interdisciplinary approach involving experts in such diverse areas as educational measurement, psychology, instructional design, statistics, and the content area. The opportunity to assemble a team that has the requisite knowledge of both assessment design and instructional design is rare. Consequently, while students could benefit from interacting with learning environments that contain well-designed embedded assessments, not enough are produced.

The primary problems of the current paper-pencil tests

Billions of dollars are spent each other on education, yet there is wide spread dissatisfaction with our educational system among educators, parents, policymakers and the business community. Efforts to reform and restructure schools have focused attention on the role of assessment in school improvement. After years of increases in the quantity of formalized testing and the consequences of inappropriate test scores, many educators have begun to doubt the measures used to monitor student performance and evaluate programs. They claim that traditional measures fail to assess significant learning outcomes and thereby undermine curriculum, instruction and policy decisions. The timed nature of the tests and their format of one right answer has led teachers to give students practice in responding to artificially short texts and selecting the best answer rather than inventing their own questions or answers. When teachers teach to traditional tests by providing daily skill instruction in formats that closely resemble tests, their instructional

IR021808

BEST COPY AVAILABLE

practices are both ineffective and potentially detrimental due to their reliance on outmoded theories of learning and instruction.

Contrary to past views of learning, cognitive psychology suggests that learning is not linear, but that it proceeds in many directions at once and at an uneven pace. Conceptual learning is not something to be delayed until a particular age or until all the basic facts have been mastered. People of all ages and ability levels constantly use and refine concepts. Furthermore, there is tremendous variety in the modes and speed with which people acquire knowledge, in the attention and memory capabilities they can apply to knowledge acquisition and performance.

According to Dietel et al (1991), Current evidence about the nature of learning makes it apparent that instruction which strongly emphasizes structured drill and practice on discrete, factual knowledge does students a major disservice. Learning isolated facts and skills is more difficult without meaningful ways to organize the information and make it easy to remember. Also, applying those skills later to solve real-world problems becomes a separate and more difficult task. Because some students have had such trouble mastering decontextualized "basics," they are rarely given the opportunity to use and develop higher-order thinking skills.

Recent studies of the integration of learning and motivation also have highlighted the importance of affective and metacognitive skills in learning. For example, poor thinkers and problem solvers differ from good ones not so much in the particular skills they possess as in their failure to use them in certain tasks. Acquisition of knowledge skills is not sufficient to make one into a competent thinker or problem solver. People also need to acquire the disposition to use the skills and strategies, as well as the knowledge of when and how to apply them (Dietel et al, 1991). These are appropriate targets of assessment. Thus, different types' alternative assessments are being considered now.

Alternative Assessment on Computer-Based Learning Environments

Alternative assessment based on cognitive modeling may be a more valid indicator of students' knowledge and abilities because they require students to actively demonstrate what they know. The following were described by Leighton et al (1999) as the characteristics of alternative assessment: (a) assessment should be integrated with instruction; (b) it should be transparent, that is, it should help students learn to monitor, self-evaluate, and reflect on their own performance; (c) assessment tasks should be authentic, extended over time, meaningful and challenging; (d) students should have access to tools, resources and coaching support during assessment of their performance and learning; (e) assessment should be diagnostic, providing information about students' knowledge, cognitive processes, misconceptions and errors during performance of problem-solving tasks; (f) students' knowledge and their ability to apply (transfer) their knowledge to new or novel problems should be assessed; (g) ability to collaborate effectively with others in solving problems should be assessed. These characteristics of alternative assessment address limitations of traditional testing approaches and promote more authentic assessments of knowledge, learning, cognitive processing, and problems-solving skill in complex natural domains and contexts of task-oriented action. And some studies show that the integration of assessment and Intelligent Tutoring System (ITS) could play a role as the alternative assessment framework as a problem-solver to cure the current malfunction ((Dietel et al, 1991). However, because designing and confirming validity of a cognitive assessment is on very early developing stage, needless to say, embedding such one on ITS that is already hard to develop should be double-efforts. Therefore, as I mentioned earlier, many researchers and practitioners for ITS tend to assess learners temporarily only for their specific purpose. Consequently, such assessment methods being used so far are not easy to transfer to another ITS. RAF would help solving such barriers.

Defining of Evidence-Centered Design (ECD) Methodology

The RAF concept is based upon the evidence-centered design methodology described in Mislevy, Steinberg, Almond, Haertel, & Penuel (2001). To grasp RAF's essential mechanism better, it will be necessary to consider what ECD is.

Its essence is focused on reasoning on the basis of evidence. When we reason from masses of data of different kinds, humans interpret complex data through some underlying "story" - a narrative, or an organizing theory. We attempt to weave a sensible and defensible story around the specifics. And it is the story that builds around what we believe to be the generative principles and patterns in the domain. Based on their story and principles, we can use them as building blocks when we tackle some novel problems that pose new questions or presents new kinds of evidence. In this spirit, RAF applies an approach structuring arguments from evidence to inference manage uncertainty (Mislevy et al, 1999). RAF approach is an evidence-centered perspective on assessment design, and object definitions and data structures for assessment elements and their interrelationships on the basis of student model, task model and evidence model.

Student Model:

The student mo del indicates, "What complex of knowledge, skills or other attributes should be assessed?" The student model variables are the terms in which we want to talk about students - the level at which we build our story, to determine evaluations, make decisions or plan instruction - but we do not get to see the values directly (Mislevy et al, 1999). The student model variables, all proficiencies that learners would acquire would be defined through cognitive task analysis. And then we just see what the students say or do and must use it as evidence about the student model variables. However, we encounter the evidentiary problem of constructing this inference from limited evidence. Hence, The inference drawn from what the learners say or do may be also on the basis of probability rules, which may be able to rule out its uncertainty.

Evidence Model:

The evidence model means, "What behaviors or performances should reveal those constructs and what is the connection?" The evidence model lays out our argument about why and how the observations in a given task situation constitute evidence about student model variables. There are two parts of evidence models, the evaluative submodel and the statistical submodel. The evaluative submodel extracts the salient features of the "work product", that is, whatever the student says, does, or creates in the task situation. The statistical submodel updates the student model in accordance with the values of these features, effectively synthesizing the evidentiary value of performances over tasks. Defining statistical submodel is beyond this study's purpose so I will not mention it any further.

Here, I will illustrate more clearly what work product means since it is a very important element mapping between the concept and data scoring. According to Mislevy et al (1999), it is a unique human production, as simple as a mark on an answer sheet, as complex as the presentation of disconfirming evidence or a series of work product represent "observable variables". It represents what is important in a performance on the basis of our belief. And these mapping between learners work product and doing assessment could be very simple or could be very complicated to require several expert's evaluation of multiple aspects. In the same context, they can be automatic, or they can require human judgment.

Task Model:

Task model means, "What tasks or situations should elicit those behaviors?" A task model provides a framework for constructing and describing the situations in which learners act and includes specifications for the environment in which the learner will say or do, or produce something. The examples could be characteristics of stimulus material, instructions, help, tools, and affordances. It also includes specifications for the work product, the form in which what the learner says, does or produces will be captured. A particular task is produced by assigning specific values to tasks model variables and providing materials that suit the specifications there given, a task thus describes particular circumstances meant to provide learners an opportunity to act in ways that produce information about what he or she knows or can do more generally. Accordingly, the task controls the level of complexity or difficulty of circumstances that learners encounter. The task itself does not describe how we should evaluate what we see. This is specified

in the evidence model. The advanced ITS is a supreme place providing the means for producing theses situations, and capturing, evaluating and communicating what learners do there.

A Solution

One solution to this problem is to develop assessment frameworks that compile the needed assessment design expertise into reusable chunks that instructional designers can instantiate as they design learning environments. In the paper, we will describe the structure of an RAF in terms of an assessment delivery model proposed by Almond, Steinberg, and Mislevy, 2001. The figure below (adapted from Almond, Steinberg, and Mislevy, 2001) provides a conceptual representation of the components of learning and assessment delivery systems.

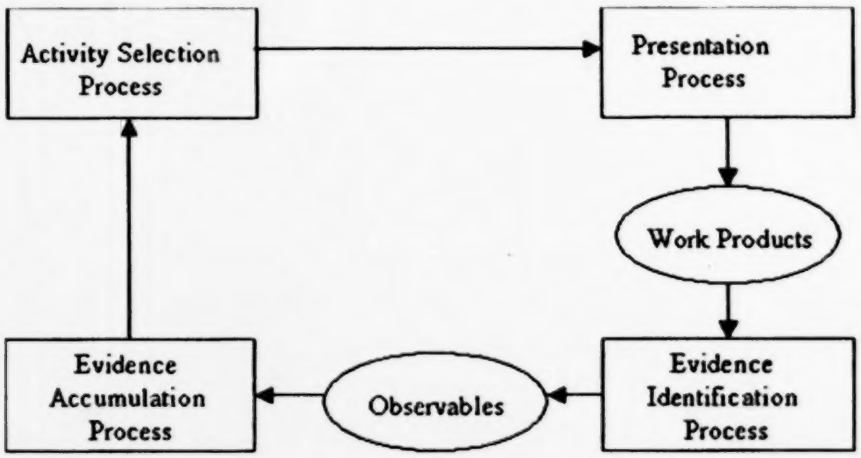

Figure 1: The Four-process model of Almond, Steinberg, and Mislevy (2001)

The activity selection process is responsible for determining on which activity a student will next work. This process could be as simple as allowing a student to choose the next activity or as complicated as an automated decision based upon the systems assessment of a students strengths and weaknesses. The presentation process is responsible for providing all direct interactions with a student including presenting all instructional and activity related material, recording all student interactions, and capturing student work. The evidence identification process scores student work, extracting evidence from the student work products as a set of "observables" – variables that contain evaluations of the student performance along a set of relevant dimensions. The evidence accumulation process uses the observables to update a student model that consists of a set of estimates of students' knowledge, skills, and abilities.

As displayed in figure 2, in an RAF several parts of this framework are defined by assessment designers before instructional designers begin their work.

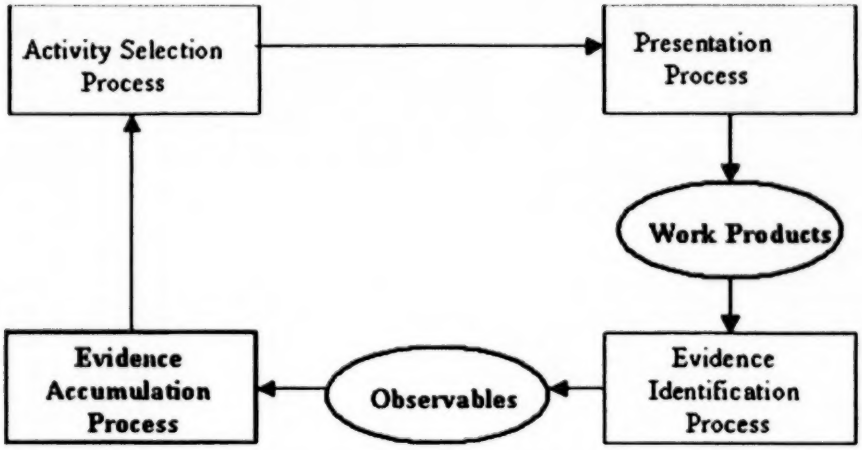

Figure 2: The Four-process model with the RAF components highlighted

In the figure, these predefined components are represented in bold text. Specifically, the evidence accumulation process is defined and implemented, and the observables and the work products are defined. In addition, the RAF includes task models as defined by Mislevy, Steinberg, and Almond (1999). The task models consist of collections of task features that include a) features that need to be present so that students will create appropriate work products (i.e. those that contain the needed evidence, b) features that affect task difficulty, and c) features that can be varied to produce task variants. The task models in the RAF define requirements for tasks that instructional designers may create. However, as will be described in the paper, considerable flexibility remains for instructional designers to design many new creative tasks as long as they incorporate selected subsets of features from the task models.

An Example: an RAF for Science Inquiry

In this section, we will describe an example RAF for the area of science inquiry. We will first present the general structure of the RAF including proficiencies, evidence, and task features. We will then provide an example of how an instructional designer might use the predefined RAF to design a computer-based learning environment with embedded assessments for middle school children. The domain of the example will be science inquiry focusing on the physics of hot air balloons. The domain of the knowledge is provided with the form of computer-based simulation program. Whenever learners encounter a challenge task, they solve it; otherwise they are not able to go to the next step. During such an itinerary, their learning progress is captured by systems, their learning outcomes are collected and measured through RAF method.

Educational Importance

An RAF is a way to represent assessment design knowledge in reusable structures. Just as the object-oriented programming movement in computer science allowed a new, deeper level of code reusability for programmers, RAFs offer instructional designers access to reusable assessment design knowledge, and hence, make it easier to for them incorporate embedded assessments in their designs, without overly constraining their creativity.

Implementation/Research perspective

A RAF will be used to as a tool to support the research project that examines the correlation between assessment and a learner's information search behavior on intelligent tutoring system (ITS). Through being used in such a project, a RAF's validity and usefulness will be examined and it can point out a RAF's limitation as well as additional usefulness in an actual performance.

References

Dietel, R.J., Herman, J.L., & Knuth, R.A. (1991). What does research say about assessment?, http://www.ncrel.org/sdrs/areas/stw_esys/4assess.htm

Mislevy, R. J., Steinberg, L.S., & Russell G.A. (2000). Evidence-centered assessment design, http://www.education.umd.edu/EDMS/mislevy/papers/ECD_overview.html

Mislevy, R.J., Steinberg, L.S., & Russell G.A. (2000). A sample assessment using the four-process framework, *the technical report*.

Steinberg, L.S., & Gitomer, D.H. (1996). Intelligent tutoring and assessment built on an understanding of a technical problem-solving task. *Instructional Science*, 24, 223-258.

Van Lehn, K. (1988). Student modeling. *Cognitively Diagnostic Assessment*. Hillsdale, NJ: Lawrence Erlbaum Associates.

BEST COPY AVAILABLE

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

NOTICE

Reproduction Basis

[X] This document is covered by a signed "Reproduction Release (Blanket)" form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.

[ ] This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").